Category Learning from Equivalence Constraints

Rubi Hammer[1,2] (rubih@alice.nc.huji.ac.il), Tomer Hertz[1,3] (tomboy@cs.huji.ac.il)
Shaul Hochstein[1,2] (shaul@vms.huji.ac.il), and Daphna Weinshall[1,3] (daphna@cs.huji.ac.il)

1. Interdisciplinary Center for Neural Computation 2. Neurobiology Department, Institute of Life Sciences
3. School of Computer Sciences and Engineering, Hebrew University, Jerusalem, 91904 Israel

Abstract

We investigated human category learning from partial information provided as equivalence constraints. Participants learned to classify stimuli on the basis of either positive or negative equivalence constraints, that is, when informed that two exemplars belong to the same category or to different categories, respectively. Knowing that in natural contexts positive constraints are usually informative while negative constraints are rarely so, we suspected that participants would not use the two types of constraints in similar ways, even in a setting in which the amount of information in the two types of constraints is identical and sufficient for perfect performance. We found that in general, people can use the two types of constraints for category learning. Further analysis revealed that when participants were provided with highly informative positive constraints, categorization performance of most participants was moderate and normally distributed. In contrast, there was a dichotomy of participants who were provided with highly informative negative constraints, with some achieving even higher performances, while that of others was significantly poorer. These results, together with those of a battery of controls, support the following conclusions: (i) People use positive constraints more intuitively, although they fail to use them perfectly. (ii) The use of negative constraints enables a less natural, but potentially more accurate categorization strategy, which many participants failed to implement even in the current simplified setting. These results are consistent with the view that people are naturally biased towards similarity-based categorization strategies (e.g. prototypes or exemplars) rather than rule-based strategies.

Introduction

It is usually assumed that people categorize objects based on their perceived similarities (Rosch & Mervis, 1975; Tversky, 1977; Medin & Schaffer, 1978; Nosofsky, 1988; Goldstone & Barsalou, 1998). Yet, since most objects have many visually-perceived features (values on physical dimensions: e.g. red color, round shape, smooth texture), similarity between objects is often difficult to define. Category learning therefore often becomes learning which object features (Tversky, 1977) or dimensions (Nosofsky, 1987) are most important for similarity judgments (Medin, Goldstone, & Gentner, 1993). In particular, different dimensions may be relevant in different domains.

Common to all category learning tasks is that they ultimately provide the classifier with clues as to the relations between particular exemplars – that two exemplars are from the same or from different categories. These relations constrain the perception and/or use of similarities (or dissimilarities) between exemplars within (or between) categories. We call a restriction that two exemplars belong to the same category, a Positive Equivalence Constraint (PEC), and a restriction that two exemplars belong to different categories, a Negative Equivalence Constraint (NEC). These two types of constraints are the building blocks of any categorization learning scenario. In particular, labeling a set of exemplars, including more than one of each of a number of categories, provides both positive and negative constraints. As an example, when a parent labels three unfamiliar animals to a young child as “a dog”, “a dog”, and “a cat”, he actually provides the child with one PEC (indicating the two dogs as belong to the same category) and two NECs (indicating that each one of the two dogs is not from the same category as the cat). Yet, the way people use these two types of constraints has not been studied directly or differentially.

Recent observations demonstrated that ecologically there are inherent differences in the properties of PECs and NECs in a multifarious world (Hertz et al, 2003). In most natural scenarios NECs are abundant since it is highly likely that randomly chosen pairs of objects belong to different categories. And yet, in most cases such pairs comprise objects that are highly different from one another in both informative and irrelevant dimensions. Thus, informative NECs, (constraints that present two highly similar objects as belonging to two different categories) are rare. On the other hand, all PECs are informative in the sense that they imply that features dimensions that differentiate between the positively paired objects are irrelevant, while the common features are candidates for being relevant for the categorization task at hand. In a study of the use of equivalence constraints in clustering and similarity-learning algorithms, it was found that generally PECs provide performance gains that are significantly better than NECs (Hertz et al, 2003). It was also found that incorporating PECs into various clustering algorithms can be straightforward and computationally feasible, while incorporating NECs is rather complicated and computationally intensive.

For these reasons, we expected that people may be more likely to adopt proper tools for the efficient integration of PECs but not NECs. Moreover, we hypothesized that this bias will remain evident even when the two types of constraints are highly and equally informative. For that it essential to examine the existence of such a bias since it might explain categorization errors as resulting from the inadequate use of all available information. To test these hypotheses, we designed an experiment in which only one or the other type of constraint was presented in each experimental condition. The amount of information provided by each type of equivalence constraint was also manipulated. First we investigated performance when the contributions of PECs and NECs to category learning simulated those expected in natural settings, by using constraints defined for *randomly* selected object pairs. We then tested performance when PECs and NECs were deliberately selected to be highly and equally informative. As a control, we tested performance when participants were provided with no equivalence constraints at all, or, at the other extreme, when provided with “meta-knowledge” – i.e. “tips” on the best strategy for the integration of constraints.

Methods

Materials

3D computer-generated pictures of "alien creature faces" were used as stimuli, as demonstrated in Figure 1. Each face was characterized by a unique combination of 5 potentially task-relevant features: chin, nose and ear shape, and skin and eye color. All (32) combinations of these 5 binary dimensions were presented in each of the 10 experimental trials. Two or three randomly selected dimensions (of the 5 possible) were relevant for category definition on each trial. Stimuli were presented on a 22-inch, high-resolution computer screen, using specially designed software.

Participants

89 university students participated in the experiment. They were randomly assigned to the different experimental or control conditions in a mixed experimental design.

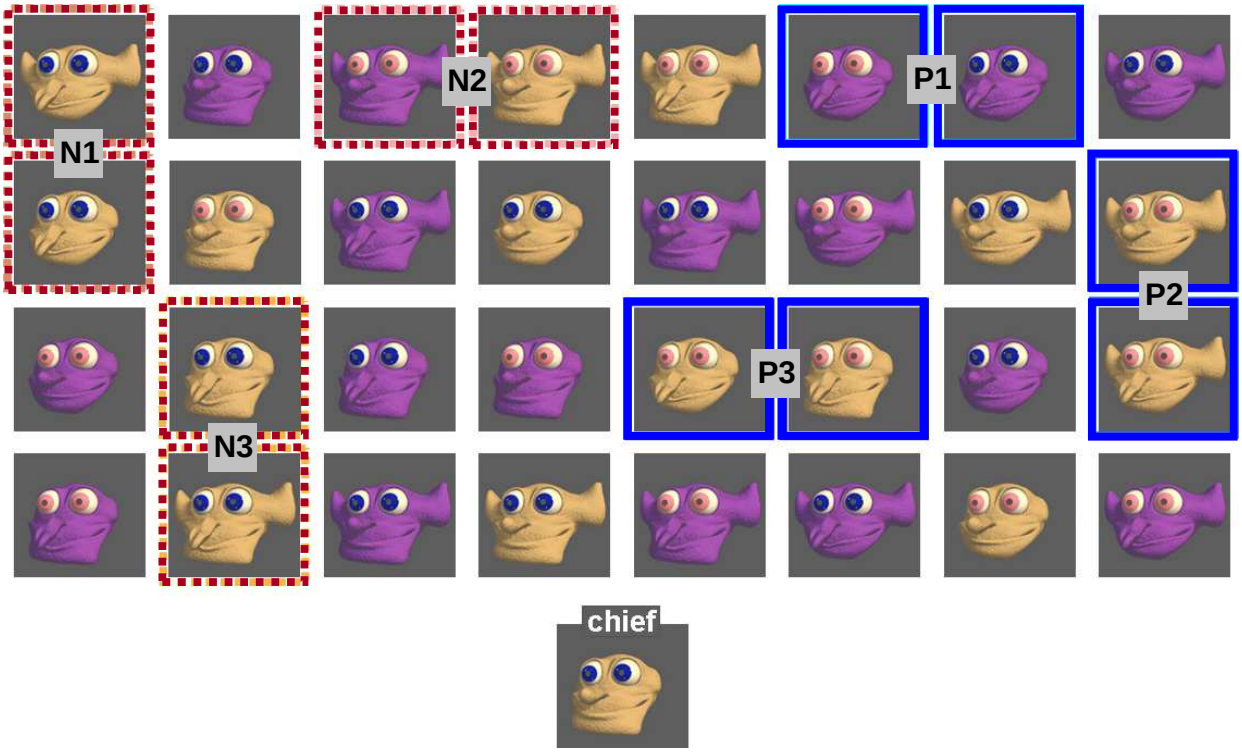

Figure 1: Example of the stimulus configuration on one specific trial. Participants decided which of the 32 test stimuli belong to the chief's tribe. Clues (constraints) were presented as frames surrounding pairs of exemplars. Positive and Negative Equivalence Constraints (PECs and NECs) are illustrated respectively as solid lines, marked P1-P3, and dashed lines, marked N1-N3. Note that in the experiment, the two types of constraints never appeared together. Highly informative constraints, as demonstrated here, present pairs of images that differ in only one feature. In the current example, participants had to learn that skin color and ear shape are relevant for categorization. Specifically, NEC N2 informs participants that skin color is a relevant dimension because it is the only dimension discriminating between the two exemplars. Similarly, N1 and N3 both imply that ear shape is relevant for categorization. P1, P2 and P3 inform participants that eye color, nose shape and chin shape are *not* relevant for categorization since these features are different in pairs that belong to the same tribe. In the highly informative constraint task, as in the current example, all the information needed for proper categorization (for either NECs or PECs, separately) was provided (see text).

Procedure

Participants were told that during the experiment they would have to learn which of the 32 "alien creatures" (test stimuli) belonged to the same tribe as the one identified as "chief" (standard). They were instructed that each trial in the experiment was independent and would necessitate learning a new way of discriminating between tribes. Participants were not informed that for each trial 2 or 3 dimensions were chosen as trial-relevant. In general, specific instructions were not given about the categorization strategy to be used for maximizing performance; rather, participants were simply asked to perform the task intuitively, using the clues provided.

Clues (constraints) were provided as colored frames around pairs of aliens, indicating that the members of the pair belong to different tribes (NEC condition) or the same tribe (PEC). Figure 1 shows examples of constraints. On each trial, 3 constraints appeared for 20 seconds together with the ensemble of alien faces. The constraints were then removed and the alien faces shuffled. Participants were then given 50 seconds to select (by drag-and-drop) those aliens that he or she thought belonged to the chief's tribe. The trial was then terminated and the next experimental trial began.

Even without using the information presented in the Equivalence Constraints, subjects could perform better than chance by simply using an associative categorization strategy that is based on some idiosyncratic similarity measure. We therefore added a control condition to establish a

performance baseline, with a "no Equivalence Constraints" (noEC) task. Twelve participants performed the same categorization task in a totally unsupervised manner, i.e. without being provided with either NECs or PECs.

After performing the noEC condition, the same 12 participants performed the randomly-selected NEC and PEC tasks (rNEC and rPEC, in counter-balanced order). The constraints were consistent with the assigned alien creature categories, but no attempt was made to select the constraints in a way that maximized the information provided for optimal performance. Note that for the reasons mentioned in the Introduction, in the rPEC condition the information provided by 3 randomly selected constraints almost always sufficed for identifying the task-relevant dimensions. This was not the case for rNECs, where the information provided was almost as poor as in the noEC task.

60 other participants were assigned to the two main experimental conditions: highly-informative Negative or Positive Equivalence Constraints (hiNEC or hiPEC). 10 additional participants performed the two conditions as a within-subject design (with counter-balanced order of performing the two conditions). The performance patterns of these ten participants were similar to those of the 60 participants in the between-experiment design; therefore, we do not address their data separately. Thus, there were 40 participants, altogether, for each of these conditions.

Highly informative constraints were pairs of test stimuli chosen so that the two images differed in only one dimension. Thus, each constraint provided information on the relevance of one dimension for tribe classification on that trial: For a hiNEC, this dimension is necessarily relevant; for a hiPEC, this dimension is irrelevant. Still, the hiPEC group could first derive the irrelevant dimensions and then infer that the rest of the dimensions were relevant. Since, on average, half of the dimensions were relevant and half were not, the amount of information provided by NECs was identical to that provided by PECs. For both groups, in each trial, the constraints were sufficient to derive explicitly all relevant dimensions needed for perfect categorization. Nevertheless, no participant performed perfectly (see results). In order that the number of constraints not be indicative of the number of relevant dimensions, we always provided 3 constraints, sometimes providing redundant information.

Performance Measures

We first report overall performance in the categorization task using the Z-score, which is a combined purity and accuracy measure defined by:

$$Z = \frac{2 * Purity * Accuracy}{Purity + Accuracy} = \frac{2 * Hits}{2 * Hits + Misses + FA}$$

where *Hits* is the number of correctly selected tribe members, *Misses* is the number of tribe members which were not selected, and *FA* (False-Alarms) is the number of incorrectly selected, non-tribe, members. Z-scores range from 0 (poor) to 1 (perfect performance). We also compared performance using the Signal Detection Theory measures, d' and criterion (Green & Swets, 1966, 1974).

Results

We present the results of three experimental conditions: In the first condition, subjects were provided with randomly selected equivalence constraints (rPECs and rNECs). As expected from the theoretical difference between these constraints discussed above, the results clearly demonstrate that randomly selected PECs are much more useful for the categorization task than randomly selected NECs. In the second – main – experimental condition, both types of constraints were designed to be highly informative (hiPEC and hiNEC). The results for this condition also demonstrate that people use PECs and NECs differently. We then present results from a third – control – experiment, which exposes additional differences between the hiPEC and hiNEC conditions.

1. Category learning from random ECs

In these experiments, subjects were provided with randomly selected equivalence constraints (rPEC or rNEC). As seen in Figure 2, participants perform considerably better when provided with rPECs than with rNECs. A set of within-subject t-tests shows that in the rPEC condition the average Z-score (0.52±0.12; mean±S.D.) was higher than in the rNEC condition (0.38±0.10), t(11)=3.46, p<0.01. Additionally, the mean Z-score in the noEC condition (0.36±0.06) was significantly lower than in the rPEC

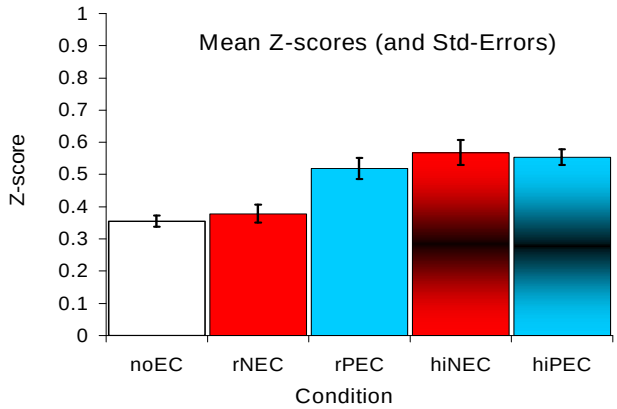

Figure 2: mean Z-score and standard error in three experimental conditions – with participants given no equivalence constraints, random constraints, or highly informative constraints.

condition, t(11)=4.28, p<0.005, but not the rNEC condition, t(11) =1.08, p=n.s.

These results are consistent with the observation stated above, that there are inherent differences in the amount of information carried in a PEC vs. a NEC. In our paradigm three randomly selected PECs almost always provided sufficient information to identify the relevant dimensions for the task, while three randomly selected NECs were almost never informative enough for fully achieving this goal. This may explain why mean performance using rNECs is very similar to that in the noEC condition.

2. Category learning from highly informative ECs

How does performance change when participants are provided with highly informative NECs (hiNEC) and PECs (hiPEC)? Recall that in these conditions, we selected the constraining pairs of stimuli so that each pair provided

information about *exactly* one relevant dimension (hiNEC), or one non-relevant dimension (hiPEC). We shall compare performance in these conditions to that with rPEC, rNEC and noEC, and then compare the results of the hiPEC and hiNEC conditions with one another, in greater detail.

As can be seen in Figure 2, mean Z-scores in both the hiNEC (0.57±0.24) and hiPEC (0.55±0.16) conditions are significantly higher than with noEC, $t(50)=4.95$, $p<0.001$ and $t(50)=6.45$, $p<0.001$, respectively. As expected, the average Z-scores in these conditions are also significantly higher than with rNEC (hiNEC: $t(50)=3.92$, $p<0.001$; hiPEC: $t(50)=3.66$, $p<0.005$), but not significantly higher than with rPEC.

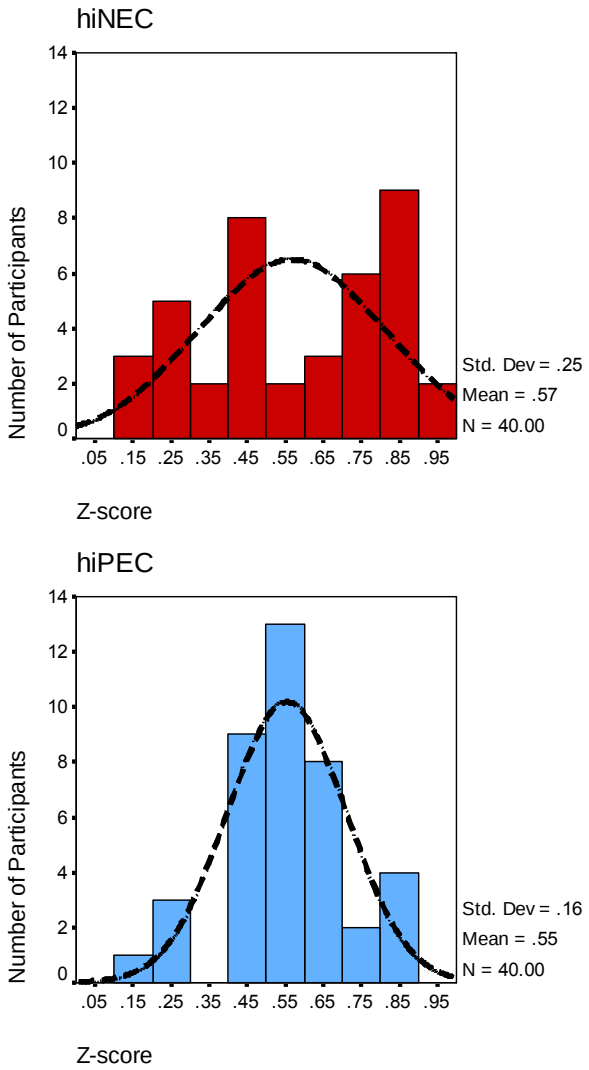

Figure 3: Observed Z-score histograms in the hiNEC (top) and hiPEC (bottom) conditions. Dashed lines represent the Gaussians defined by the Means and Standard Deviations.

At first glance, it may appear that when NECs are designed to provide the same amount of information as PECs, they both lead to similar performance. However, as shown in Figure 3, although there was no significant difference between the *average* performance in the hiNEC and hiPEC conditions, there was a highly significant difference between their standard deviations (Levene's test of homogeneity of variances: $F(78)=17.31$, $p<0.001$). Furthermore, the Shapiro-Wilk test of normality revealed that the Z-score distribution was normal in the hiPEC condition, $W(40)=0.955$ $p=0.20$, but not in the hiNEC condition $W(40)=0.932$ $p<0.05$.

These findings suggest that even under highly informative constraint conditions, in which the information provided by NECs is identical to that provided by PECs, participants use these constraints differently. In the hiPEC condition, most of the participants did not use all of the information provided by the constraints, which was always sufficient to perform the task perfectly. For this reason, hiPEC performance was usually moderate, and it was normally distributed. In contrast, in the hiNEC condition, performance was either poor or excellent. It seems that many participants did not use hiNECs correctly, but those who did obtained almost perfect performance. This suggests that NECs are less intuitive as a source of information for category learning tasks.

To analyze further the differences between use of hiNEC and hiPEC, we compared participant d' and criterion in these conditions. We found no significant difference between the average d' in hiNEC (1.87±1.10) and hiPEC (1.64±0.64) $t(78)=1.157$, p=n.s. However, the standard deviations of the d' in the two conditions did differ significantly $F(78)=19.41$, $p<0.001$. Additionally, the mean criterion used in the hiNEC condition (1.65±0.49) was significantly higher than in the hiPEC condition (1.27±0.35), $t(78)=4.03$, $p<0.001$, and their standard deviations also differed significantly, $F(78)=10.91$, $p<0.005$.

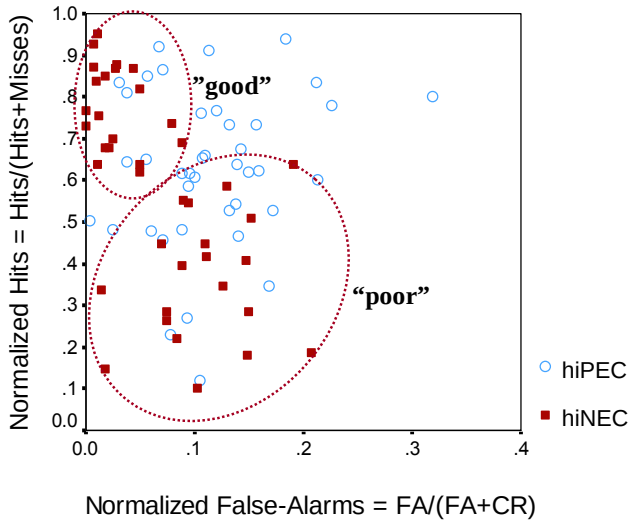

Figure 4: Participant performance in the hiPEC (empty/light circles) and hiNEC (filled/dark squares) conditions plotted on a ROC (Receiver Operating Characteristic) diagram. The dashed ellipses separate the good hiNEC performers from the poor hiNEC performers.

As seen in Figure 4, this difference in criterion leads to a higher number of False-Alarms (FAs) with hiPEC than with hiNEC. We found a highly significant negative correlation between Hits and FAs in the hiNEC condition, $r(40)=-0.59$, $p<0.001$, but not in the hiPEC condition $r(40)=0.16$, p=n.s. The above-noted highly variable performances in the hiNEC condition, together with the Hit/FA correlation, suggest that this group of participants may in fact represent two distinct subgroups: one with high Hit and low FA rates ("good" in Figure 4) and the other ("poor") with low Hit and high FA rates. This division was confirmed using the K-means

algorithm to cluster the hiNEC group into two subgroups (based on their Z-scores), marked by dashed ellipses in Figure 4. There is no significant Hit-FA correlation for either subgroup, individually, further justifying the division into separate subgroups. We denote the subgroups hiNEC-good and hiNEC-poor, respectively.

Comparing these subgroups, we found that the mean Z-score in the hiNEC-good subgroup (0.78±0.10) was significantly higher than in all other experimental conditions (e.g. compared to hiPEC: $t(58)=5.98$, $p<0.001$). On the other hand, the mean Z-score for the hiNEC-poor subgroup (0.35±0.13) was not significantly different from that with noEC, $t(30)=0.60$, p=n.s., or rNEC $t(30)=0.07$, p=n.s. (see Figure 5). In practice, participants in the hiNEC-poor subgroup behave as if they were not provided with any equivalence constraints.

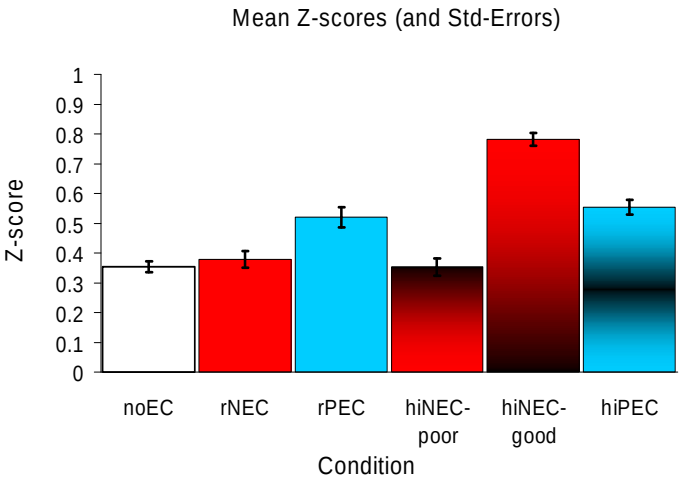

Figure 5: mean Z-score and standard error in the hiNEC-good and hiNEC-poor conditions compared to the other experimental conditions.

Taken together, these findings unravel some of the basic differences between the way that PECs and NECs are used. While hiPECs intuitively provide most people with helpful information for category learning, hiNECs provide little information to some, and enough information for almost perfect categorization performance to others.

3. Category learning from highly informative ECs together with "Meta-Knowledge"

If indeed most people can effectively use PECs in an intuitive manner, while using NECs requires specific expertise obtained by only some people, we may expect that teaching participants how to use hiNECs will substantially increase their performance, while teaching them how to use hiPECs will not provide as much benefit. To check this hypothesis, we tested 7 additional participants (using a within-participant design) with exactly the same hiPEC and hiNEC conditions, augmented with additional "Meta-Knowledge" – guidelines regarding the best strategy for integration of constraints. Specifically, before performing the hiPEC condition, participants were informed that they should exclude the dimension discriminating between each two constrained exemplars (and reserve judgment about the rest), since this dimension is irrelevant for the categorization task. Before performing the hiNEC condition, the same participants were informed that they should take into account the dimension discriminating between each two constrained exemplars because it is relevant for the categorization task.

As seen in Figure 6, "meta-knowledge" was extremely helpful in improving participant performance in the hiNEC condition but not the hiPEC condition. Performance in the directed hiNEC condition (0.90±0.06) was significantly higher than in the original non-directed hiNEC condition (0.57±0.24) $t(45)=7.45$, $p<0.001$. Performance in the directed hiNEC condition was also higher than in the hiNEC-good subgroup alone (0.78±0.10). In contrast, performance in the directed hiPEC condition (0.63±0.21) was not significantly higher than in the non-directed hiPEC condition (0.55±0.16) $t(45)=1.20$, p=n.s. These findings indicate that participants provided with PECs naturally operate a categorization strategy that leads to satisfactory performance equal to that

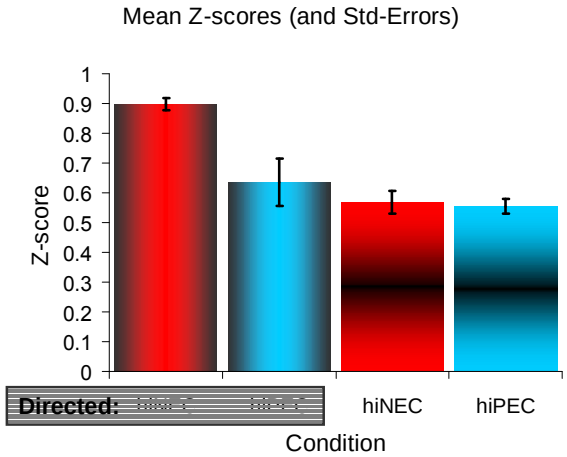

Figure 6: mean Z-score and standard error of the directed hiPEC and hiNEC conditions compared to the non-directed hiPEC and hiNEC conditions. As can be seen, directions helped participants only in the hiNEC condition.

achievable when provided with best-strategy "tips".

The moderate performance even in the directed hiPEC condition may be explained by a possibly inherent disadvantage of PECs, amplified in our experimental paradigm: PECs, unlike NECs, directly specify an irrelevant dimension. Thus, participants can only learn indirectly which dimensions are task-relevant. Since participants are not aware of all the *potentially* task-relevant dimensions, it is very likely that they will miss relevant dimensions even when correctly filtering out the irrelevant dimensions by using an optimal strategy. Missing or disregarding the relevance of a dimension would lead to more False-Alarms, and indeed, participants in the PEC condition have higher False-Alarm rates, as shown in Figure 4.

Discussion

The goal of the current study was to investigate mechanisms underlying categorization, by comparing performance when using Positive Equivalence Constraints (PECs) vs. Negative Equivalence Constraints (NECs). Due to the inherent differences between PECs and NECs, we expected that people naturally develop more effective tools for the efficient integration of informative PECs than NECs. In fact, we hypothesized that this bias may be also evident when the two

types of constraints are highly and equally informative. Our findings confirm these hypotheses, indicating that most people have an inherent or early-acquired mechanism for deriving useful information from PECs, but not from NECs.

What strategies are used with PECs?

We report the somewhat surprising result, that performance with PECs derives little benefit from the availability of meta-knowledge about the optimal strategy to perform the task. This may be explained by one of the following: (i) participants are already using the optimal strategy, and so the "tips" give them no additional information, (ii) participants' default strategy, although different, leads to similar performance levels as the optimal strategy, or (iii) the default strategy is so natural and intuitive, that participants are reluctant to shift to a potentially better strategy.

What is the default strategy that people use with PECs? It seems that PECs are naturally suited to an exemplar-like strategy, based on the storage of a large number of examples, or to a prototype-like strategy, based on the abstraction of typical elements in the class. In our setup, however, there are no explicit classes, and there is little meaning to the notion of class exemplars or prototypes. This does not rule out the possibility that people still use these kinds of strategies, trying to infer (or guess) class examples or class prototypes from the constraints provided.

Recall that the only information that participants can reliably derive from PECs is the identity of the dimensions that are relevant or irrelevant. Thus, another strategy may be to commence with a guess of the set of all potentially relevant dimensions (beginning perhaps with the most salient dimensions, corresponding to features that are common to the first pair of exemplars seen). Afterwards, one rules out dimensions sequentially, as evidence is accumulated that some dimensions are irrelevant. Participants who use this strategy in the hiPEC condition may miss relevant (less salient) dimensions, resulting in many False Alarms.

A third alternative strategy is the one offered to participants in the directed hiPEC experimental condition as additional "meta-knowledge". This strategy, if used correctly, should lead to perfect performance. Specifically, for each pair in a PEC, find the single dimension which differentiates between the two examples, and identify this dimension as irrelevant. After all the irrelevant dimensions are identified, find the set of relevant dimensions (by performing a set complement operation). As in the preceding strategy, the use of this strategy may result in elevated False-Alarms, since participants may have difficulty inferring the set of all possible relevant dimensions. Similarly, in real-world cases, the full group of possible dimensions may not be known or even inferable – in which case using the optimal strategy for PECs will not guarantee perfect performance.

PECs vs. NECs

When not provided with additional instructions, highly informative NECs were effectively used by only some participants, suggesting that this source of information requires a less intuitive skill. Moreover, performance with NECs, in contrast to PECs, benefited significantly from meta-knowledge about the optimal categorization strategy. Two interpretations can explain this difference: (i) Participants were more open to advice on how to use NECs because they did not have a strong intuitive idea of what to do *a priori*; (ii) it was easier for the participants to learn the optimal strategy with NECs, perhaps because it does not involve a set complement operation. Note that the performance of the subgroup of good-hiNEC participants was about as good as that of those provided with meta-knowledge, suggesting that this subgroup was able to use a similar strategy even without instruction.

Our findings point out one way in which the human psyche reflects statistical properties of objects and categories in the world – most people have early-adopted tools that are useful for integrating PECs since PECs are less common but always informative. This innate strategy is good enough so that even when being explicitly provided with a better categorization strategy for using PECs, performance is not significantly improved. In the case of NECs, only some people have a useful strategy for proper use of this source of information. When provided with guidelines for the best categorization strategy using hiNECs, performance significantly improved, further indicating that there is no inherent strategy for optimally integrating NECs for categorization.

The implications of the current findings may be crucial for understanding known phenomena in category learning, and they may provide an effective tool for predicting performance in different category learning tasks. As an example, the tendency of children to over-generalize when classifying objects (Neisser, 1987) may be seen as a consequence of using mostly PECs, which, as pointed out above, can lead to disregarding relevant dimensions and a subsequent higher rate of False-Alarms. Only later in life, is over-generalization reduced when more refined strategies are acquired, such as the use of rare, but informative, NECs.

References

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition, 65*, 231-262.

Green, D. M. & Swets, J. A. (1966 and 1974). *Signal Detection Theory and Psychophysics*. New York: Wiley and Huntington NY: Krieger

Hertz, T., Shental, N., Bar-Hillel, A., & Weinshall, D. *Enhancing Image and Video Retrieval: Learning Via Equivalence Constraints.* IEEE Conf. on Computer Vision & Pattern Recognition, Madison WI, June 2003

Medin, D. L., & Schaffer, M. M. (1978). Context Theory of Classification Learning. *Psychological Review*, 85, 207-238.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respect for similarity. *Psychological Review*, 100(2), 254-278.

Murphy, G. & Medin, D. L., (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.

Neisser, U., ed. (1987) *Concepts and Conceptual Development* Cambridge MA Cambridge U. Press:.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 87-108.

Rosch, E., & Mervis, C. D. (1975). Family resemblance studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: Detection, search, and attention. *Psychological Review*, 84, 1-66.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327-352.